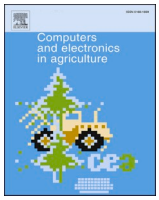
Computers and Electronics in Agriculture 224 (2024) 109209

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



Hyperspectral imaging predicts macadamia nut-in-shell and kernel moisture using machine vision and learning tools

Michael B. Farrar [a,*], Reza Omidvar [b], Joel Nichols [a], Daniele Pelliccia [c], Suhad Lateef Al-Khafaji [d], Iman Tahmasbian [a,e], Nimanie Hapuarachchi [a], Shahla Hosseini Bai [a]

[a] Centre for Planetary Health and Food Security, School of Environment and Science, Griffith University, Nathan, QLD, 4111, Australia
[b] School of Engineering and Built Environment, Griffith University, Nathan, QLD, 4111, Australia
[c] Instruments & Data Tools Pty Ltd, Rowville, VIC, 3178, Australia
[d] School of Information and Communication Technology, Griffith University, Nathan, 4111, Australia
[e] Department of Agriculture and Fisheries, Queensland Government, Toowoomba, QLD, 4350, Australia

ARTICLE INFO

Keywords:
Automation
Food quality control
Macadamia nuts
Moisture concentration
Post-harvest quality
VNIR spectroscopy

ABSTRACT

Tree nuts are a convenient and nutritious food source and recently considerable attention has been placed on quality assessment to provide high quality nuts and improve consumer satisfaction. Moisture is a critical parameter for tree nut quality and is routinely monitored throughout post-harvest processing. However, current direct methods to assess nut moisture are based on using limited numbers of representative sub-sets and are destructive. This study aimed to use hyperspectral imaging and machine learning (ML) to predict moisture of individual macadamia nuts during post-harvest processing. Specifically, we aimed to compare data extraction methods (automatic *vs.* manual) and nut orientation (base-up, base-down and combined orientations) during imaging in predicting moisture for nut-in-shell and kernels. We also explored minimum wavelength numbers to predict moisture. Spectra were obtained from images of nuts in two orientations and extracted using manual and automatic methods prior to development of partial least squares (PLSR), artificial neural network (ANN), support vector machine (SVM) and Gaussian process regression (GPR) models. Kernel moisture prediction was more accurate using automatically extracted spectra, whereas nut-in-shell moisture prediction accuracy was similar for either method. For kernels, combining the spectra from two images of nuts in base-up and base-down orientations provided similar prediction accuracy ($RMSE_T = 0.308$ %), compared with spectra from one image ($RMSE_T \geq 0.341$ %), and for nut-in-shell, using spectra from one image also provided similar accuracy ($RMSE_T \approx 1.2$ %) as using both images combined. PLSR models predicted moisture with very high accuracy for both nut-in-shell ($R^2_T = 0.96$, $RMSE_T = 1.20$ %, RPD = 5.15) and kernels ($R^2_T = 0.99$, $RMSE_T = 0.308$ %, RPD = 11.05) following selection of ten important wavelength bands between 760 and 967 nm. ANN and GPR also achieved equivalent ($R^2_T = 0.99$) highest accuracy predictions for kernels, however, all wavelengths were required, which would increase computational processing time for high volume applications. The important wavelength bands required to develop accurate models for macadamia moisture prediction are consistent with other food and nut products and prediction accuracies are possible for process control applications using only 10 wavelength bands. Several ML models including PLSR, ANN and GPR are suitable for use with Vis/NIR hyperspectral images to predict macadamia moisture, however, for industrial applications where high volume through-put is required, using PLSR with limited selected wavelength bands is recommended. Overall, hyperspectral imaging combined with computer vision software and ML models showed significant potential to predict moisture concentration of macadamia during post-harvest processing.

1. Introduction

Tree nuts are a food source high in protein, nutrients, fibre and healthy fats (Maestri, 2023; Ros, 2010). Tree nuts contain unsaturated fats and antioxidants giving them wide-ranging metabolic and cardiovascular benefits when consumed as food (Buthelezi et al., 2019;

* Corresponding author.
E-mail address: m.farrar@griffith.edu.au (M.B. Farrar).

https://doi.org/10.1016/j.compag.2024.109209
Received 8 March 2024; Received in revised form 4 June 2024; Accepted 25 June 2024
Available online 3 July 2024

Maestri, 2023; Ros, 2010; Yang et al., 2009). However, tree nuts are susceptible to a range of quality defects and effective post-harvest processing and handling is critical to maximise shelf life and nutritional quality for consumers (Gama et al., 2018). For example, high moisture concentration and humidity can induce mould and microbial growth, rancidity and internal discoloration of nuts and decrease nutrients overtime leading to lower quality product and consumer dissatisfaction (Gama et al., 2018; Kowitz and Mason, 2003; Walton et al., 2013). Nut moisture varies among batches collected at different farms and nuts can also reabsorb moisture from the atmosphere, therefore, moisture must be continually monitored throughout the post-harvest supply chain to maintain high quality product (Walton and Wallace, 2008). Commonly used direct methods to quantify nut moisture include gravimetric, coulometric or chemical such as the Karl Fisher titration (Zambrano et al., 2019). However, direct methods are destructive, and therefore, the same samples cannot be used for further analysis, sale or consumption (Zambrano et al., 2019). Moreover, current direct methods are only suitable for testing a small number of samples that forms an assessment of a representative sub-set of any given batch and is therefore subject to sampling biases (Buthelezi et al., 2019; Zambrano et al., 2019). Hence, rapid and non-destructive assessment methods are being sought.

Hyperspectral imaging (HSI) is emerging as a versatile and effective tool to non-destructively determine a wide range of physical and chemical parameters in food and agricultural applications (Dung et al., 2023; Farrar et al., 2021; Farrar et al., 2023; Hapuarachchi et al., 2023; Malmir et al., 2020). Most commonly, the visible and near-infrared (Vis/NIR) and short-wave infrared (SWIR) are being investigated for plant and food applications because shorter NIR wavelengths afford increased penetration depth for object analysis (Adebayo et al., 2016; Manley, 2014). The Vis/NIR spectral range between 400 and 1000 nm contains wavelength bands recorded by common red, green, blue (RGB) vision systems (Manley, 2014). However, in addition to RGB wavelength bands, hyperspectral images contain additional information from the infrared region (Manley, 2014). Hyperspectral images have been used to assess moisture concentration of complex and heterogeneous food materials such as red meats, prawns, soybeans and leaves (Huang et al., 2014; Kamruzzaman et al., 2016; Sun et al., 2019; Wei et al., 2019; Wu et al., 2012). However, moisture concentration of macadamia tree nuts has not been previously determined using HSI a system.

Hyperspectral images must be processed prior to predictive model development to differentiate and separate objects with varying shapes from the image background. This process may be done manually in scoping studies but needs to be automated using computer vision for real world applications (Ma et al., 2021; Wu et al., 2012). Generally, many nuts are presented in a single hyperspectral image that includes both target nuts and the background (Wu et al., 2012). Moreover, the shape of nuts is irregular, their chemical composition is heterogeneous and different parts of a given nut may have diverse chemical compositions (Wallace and Walton, 2011). For example, using spectra extracted from different spatial regions of interest (ROI), and the orientation of tree fruits, such as avocado, during imaging can affect the prediction accuracy of developed models (Kämper et al., 2020). However, object orientation and extraction method must still be examined for macadamia nuts which are much smaller than tree fruits like avocado. Spectral data can be extracted from images manually (handcrafted) or using computer vision software to automatically identify nuts using image processing tools including boundary detection (Ma et al., 2021; Wu et al., 2012). For example, boundary detection can effectively outline and detect objects in hyperspectral images by exploiting the near-infrared or short-wave infrared spectrum (Al-Khafaji et al., 2021). Currently, there is no definitive agreement within the literature regarding the optimal way to image and extract hyperspectral data from tree nuts for use in moisture prediction.

Hyperspectral imaging systems generate very large datasets that often contain highly related and analogous spectral information making their interpretation difficult using traditional statistical methods (ElMasry et al., 2013). Machine learning (ML) techniques are suitable for processing large and complex datasets and have also been shown to improve prediction accuracy for some datasets (Saha and Manickavasagan, 2021). Until recently, ML techniques were considered too computationally demanding for high throughput systems (Saha and Manickavasagan, 2021), however, rapid advancement in computer processing efficiency increased the potential for ML to be deployed in agricultural and food processing applications (Saha and Manickavasagan, 2021). Machine learning includes algorithms that can learn from data without relying on explicit programming (Saha and Manickavasagan, 2021). Supervised learning is the process of training a model on labelled training data (reference data) to make classification or prediction about the future (independent test) data (Saha and Manickavasagan, 2021). Supervised methods include, amongst other, Artificial Neural Networks (ANN), Gaussian Process Regression (GPR), Partial Least Square Regression (PLSR) and Support Vector Machines (SVM) (Saha and Manickavasagan, 2021). Additionally, the large quantity of data captured within hyperspectral images can be pre-processed to reduce spectral noise and unimportant information and post-processing may also be implemented for selection of key wavelength bands to reduce redundancy of data and improve processing speed (Liu et al., 2014; Saha and Manickavasagan, 2021). Different hyperspectral sensors operate within different spectral regions and because the structure of each target object varies, each specific application must also be systematically investigated to find the appropriate wavelengths for optical assessment (Adebayo et al., 2016). Important wavelength bands can be selected with a variety of methods including selection of large regression coefficients and selection of variables to minimise redundancy in the dataset (Liu et al., 2014; Peng et al., 2005). The capability of a range of supervised ML models in combination with a variety of pre- and post-processing techniques is required to be examined to predict moisture concentration of macadamia tree nuts.

Macadamia is a rapidly expanding tree nut crop and global production and consumption have more than doubled during the last decade (INC, 2023). Macadamia industries are now well established in South Africa, Australia, China, Brazil and Kenya and industries are developing in Guatemala, Vietnam, Malawi and India (INC, 2023). Like other tree nuts, macadamia nuts can be stored and exported as nut-in-shell or kernel and this decision can affect their shelf life and quality for end consumers (Gama et al., 2018). Moisture is a critical parameter during macadamia harvesting and processing, and quantitation is required at several points during post-harvest processing (Walton et al., 2013). Firstly, nuts fall to the ground and are de-husked on-farm at approximately 25 % moisture concentration (Walton and Wallace, 2011). Secondly, growers are paid based on sound kernel recovery (not including shell) adjusted by moisture of nut-in-shell at time of delivery, and therefore, moisture must be determined for each consignment on arrival to a processor (Guthrie et al., 2004). Nut-in-shell is then dried to ≈3% moisture prior to cracking, and therefore, regular assessment of moisture for each drying bin is necessary to inform decisions about when to crack. Finally, kernels are packaged at ≈1.5 % moisture and are tested again prior to packaging to ensure quality control (Walton and Wallace, 2011). The current gravimetric laboratory method for determining nut-in-shell and kernel moisture involves oven drying nuts for 24 hrs, and therefore, moisture data for decision making lags sample collection by at least one day (AMS, 2022). Development of a method using HSI that can rapidly determine moisture concentration will facilitate real-time sorting based on moisture during post-harvest processing allowing for agile decision making to improve efficiency.

This study aimed to develop a tool for accurate assessment of macadamia nut-in-shell and kernel moisture concentration for use during post-harvest processing to enable more timely management decisions. Macadamia kernels generally display two distinct hemispheres: 1) a flattened/spherical shaped ‘base’ that forms within the hilum end of the shell, and 2) a pronounced ‘peak’, ‘crest’ or ‘shoulder’ that forms on the

micropylar half adjacent to the attachment point (micropyle) with the shell (AMS, 2021; Wallace and Walton, 2011; Walton and Wallace, 2008; Walton and Wallace, 2012). Dropped kernels often settle on a flat surface in two orientations: 1) on their crest with the base facing upward (base-up); or 2) on their base with the crest facing upward (crest-up). It remains uncertain how collecting spectral data from nuts imaged in different orientations and how that data is extracted will affect prediction accuracy of various ML models. Therefore, we investigated the: 1) efficacy of using computer vision to automatically extract hyperspectral data from images for use in model development, 2) effect of object orientation during imaging on prediction accuracy and using a pooled dataset combined from multiple image orientations, and 3) most effective machine learning models and selection of important wavelengths for predicting moisture concentration of macadamia nut-in-shell and kernels. Development of a rapid method that can assess nut moisture non-destructively and for individual nuts would represent a significant step forward for quality control in post-harvest processing.

2. Methods

2.1. Experimental design and sample collection

Macadamia nut-in-shell and whole kernel samples were collected from commercial processors located in two states, Queensland and New South Wales, Australia. We aimed to collect samples to represent a wide range of varieties, growing conditions, and moisture concentrations (Farrar et al., 2021). Therefore, samples were collected in May, July and September during the 2022 season to ensure representation of early, middle and late harvest varieties and variation introduced by different grower practices and locations. Firstly, nut-in-shell samples were

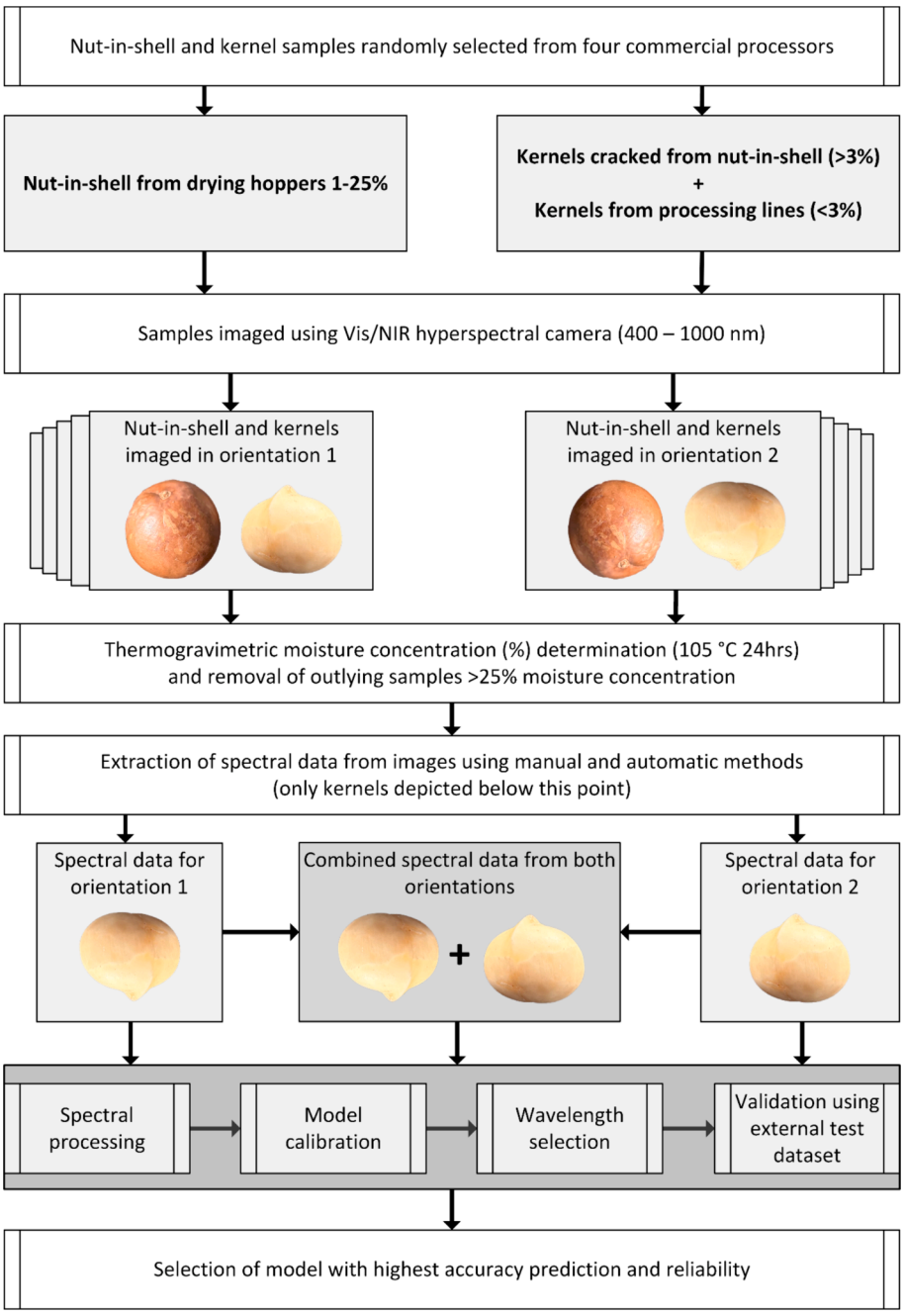


Fig. 1. Flowchart depicting the experimental design and process of model development and evaluation.

collected from drying hoppers at different drying stages where approximate moisture concentration varied between 1 % and 25 % (Fig. 1). In total, thirty ≈1 kg bags of nut-in-shell macadamias were collected, and 15 individual nuts were randomly sub-sampled from these bags for imaging.

Kernel samples were collected in two stages. The first stage involved collecting kernels directly from processing lines following mechanical cracking by laboratory staff at the processing facility and at < 3 % moisture. In total, thirty ≈150 g bags of kernels were collected, and 15 individual kernels were randomly sub-sampled from these bags. The second stage involved manually cracking kernels from collected nut-in-shell samples (described earlier) with moisture concentration between 3 % and 25 % immediately prior to imaging (Fig. 1). The second stage of kernel sampling was performed to deliberately obtain kernels with moisture concentration > 3 %. All samples were double bagged during transport and stored at room temperature (≈25 °C) to minimise moisture exchange with the atmosphere or between bags of nuts with different moistures prior to hyperspectral imaging and moisture determination within two weeks of collection. In total, 485 nut-in-shell and 529 kernels were imaged and analysed for moisture concentration (%) in this study (Supplementary Table S1).

2.2. Hyperspectral imaging, image processing, and analysis of macadamia reflectance spectra

2.2.1. Acquisition of hyperspectral images

Hyperspectral images were taken using a line scanning hyperspectral camera (Resonon Pika XC2, Bozeman, MT, USA). The hyperspectral camera had a 23.1° field of view, and spectral range of 400–1000 nm with 1.3 nm spectral sampling intervals to total 462 spectral wavelength bands. Calibration of the instrument was completed prior to image acquisition. Dark (D) current noise was obtained by taking an image with the lens cap on and white (W) correction was obtained by taking an image of highly reflective Lambertian material providing 99 % reflectance. Reflectance (R) was calculated from raw spectral reflectance (I_0) using Eq. (1):

$$R = (I_0 - D)/(W - D) \quad (1)$$

Two unique images of nut-in-shell and kernel samples were obtained during image acquisition. Samples were placed on a custom-made plywood tray to eliminate movement during imaging. The tray had recesses (≈10 mm × 5 mm) drilled to locate macadamia samples and was painted black with silicone-based paint to reduce light scattering (White Knight, Hi temp paint). First, nut-in-shell samples were placed in a random position on the tray and imaged (Fig. 2a,b), then rotated ≈180° and re-imaged. For kernel samples, images were acquired with the nut placed on the tray in two unique orientations: 1) base-up (Fig. 3a,b) and 2) base-down (Fig. 3c,d) orientations. Instrument calibration and image acquisition were conducted using SpectrononPro™ software (v3.4.6, Resonon, Bozeman, MT, USA).

2.2.2. Selection of regions of interest (ROIs) using manual and automatic methods

Spectral data pertaining to each individual nut sample were extracted from hyperspectral images using two methods: 1) manual; and 2) automatic using computer vision software. Firstly, spectral data was extracted manually (handcrafted) using the rectangle tool to highlight individual nut ROIs within the SpectrononPro™ software developed by the instrument manufacturer (Fig. 2a; Fig. 3a,c) (v3.4.6, Resonon, Bozeman, MT, USA). Secondly, spectral data was extracted from nut ROIs using a computer vision algorithm developed with *OpenCV* (v4.7.0) library in Python language for performing image analysis (Fig. 2b; Fig. 3b, d) (Bradski, 2000). In this work, we adapted the contour detection image processing algorithm to detect the nut boundary and retrieve the nuts' centre point features. Contour detection encodes the boundaries of an object contained within an image, and is useful for various applications in image processing, including segmentation and object detection (Park et al., 2022). Contour detection was applied using one selected wavelength (900 nm wavelength for nut-in-shell and 518.98 nm for kernels, respectively) to detect all nuts in any given image (Fig. 2b; Fig. 3b,d). First, the image wavelength was converted to a greyscale image, and then binary thresholding followed by dilation was used to clear the image background and remove noise (small holes). Second, we applied the contour detection algorithm to identify the boundaries of the foreground nuts, then the area of the contour was

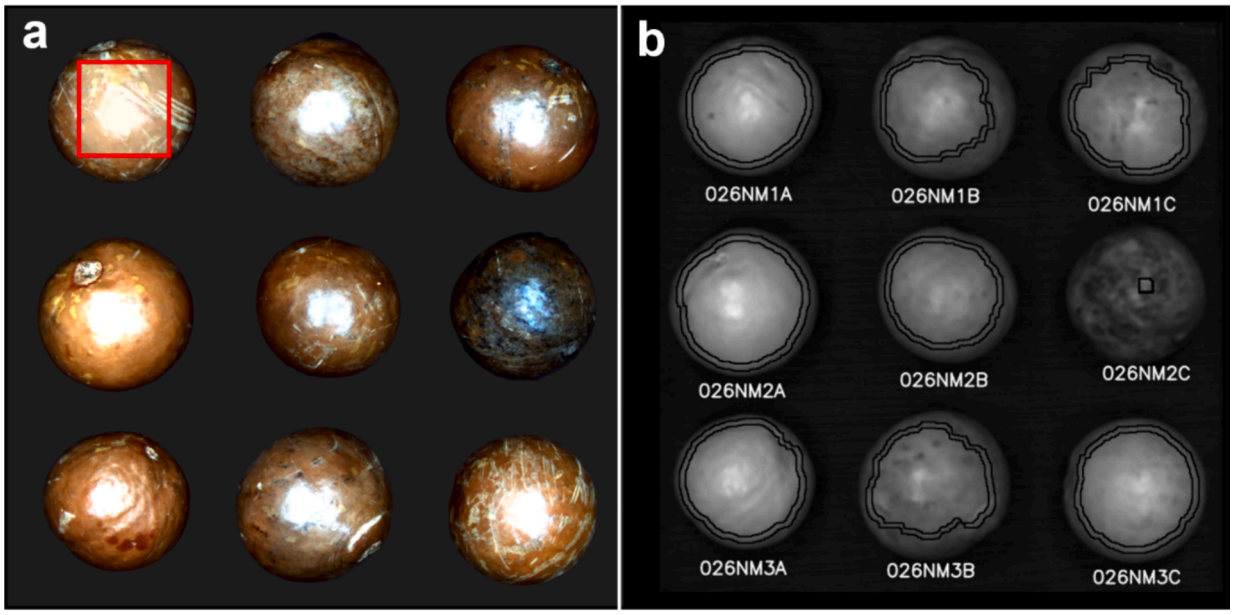


Fig. 2. **(a)** Pseudo RGB hyperspectral image of macadamia nut-in-shell, the red box represents an example of a manually selected region of interest (ROI) for an individual nut-in-shell containing the pixels from which mean spectral data were extracted and used in model development; **(b)** greyscale image of the same nuts depicted in (a) selected using computer vision, the black circles represent the inner and outer contours detected and the inner circle represents the ROI containing the pixels from which mean spectra were extracted and used in model development.

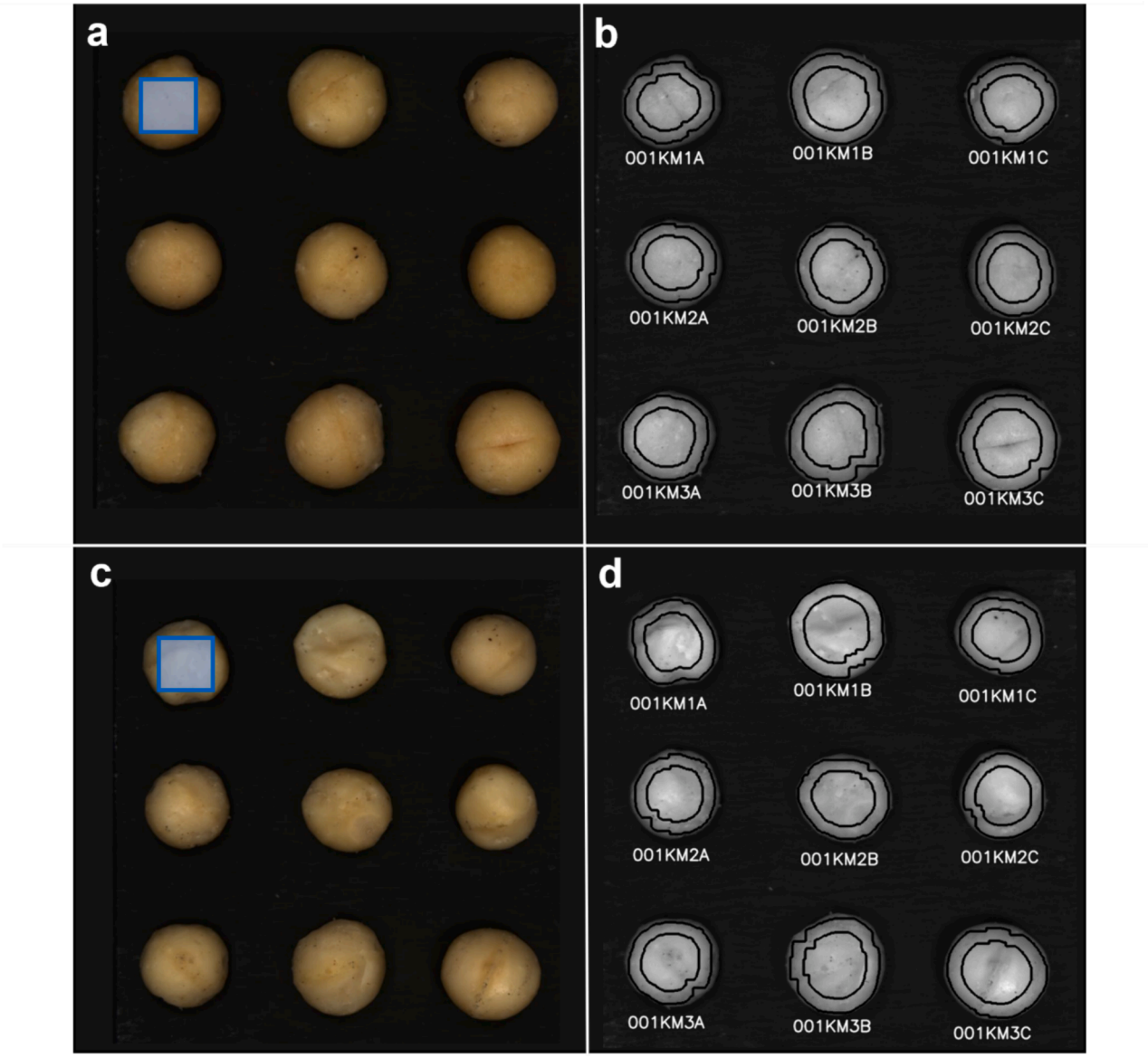


Fig. 3. (a) Pseudo RGB hyperspectral image of macadamia kernels in base-up orientation; **(b)** greyscale image of the same nuts depicted in (a) selected using computer vision; **(c)** pseudo RGB hyperspectral image of macadamia kernels in base-down orientation. The red boxes in (a) and (c) represent examples of manually selected regions of interest (ROIs) for the individual kernels containing the pixels from which mean spectra were extracted and used in model development; and the black circles in (b) and (d) represent the contours detected using computer vision for each individual kernel and the inner circle represents the ROI from which mean spectra were extracted and used in model development.

extracted as an ROI. Then, we calculate the average of the spatial points of each detected area (nut) at each spectral wavelength to produce one spectral vector of length B, where B is the number of wavelengths (462). As a result, a spectral matrix of dimension N × B were obtained, where N is the number of the nuts extracted from the image, and B is the number of spectral wavelengths in the hyperspectral image. For samples that were not detected, or incorrectly detected using computer vision, an elliptical ROI representative of the contour detection algorithm was manually handcrafted and the extracted spectral data was used to substitute the automatically extracted for that given sample (Table S1, Supplementary Fig. S1). Substitution of spectra for problematic samples was undertaken to balance the datasets prior to model development and ensure representation of problematic samples in both the manual and automatically extracted datasets allowing for future comparison of models developed during the study.

Finally, mean reflectance for both hyperspectral image datasets of nut-in-shell and kernels was calculated using data extracted from the two unique images and used as combined image data in further model development. In total, we extracted spectral reflectance data from 485 nut-in-shell and 529 kernel samples for use in model development and evaluation during this study, and spectral data substitution occurred in 3.8 % of total cases for samples that presented with surface irregularities that were problematic for automatic detection (Supplementary Table S1).

2.3. Determination of macadamia moisture concentration

Moisture concentration of nut-in-shell and kernel samples was determined using the Australian Macadamia Society (AMS) drying method for kernel recovery determination (AMS, 2022). Specifically, samples were weighed and then dried at 105 °C for 24 hrs in individual tin foil trays placed in a dehydrating laboratory oven (TO-150G,

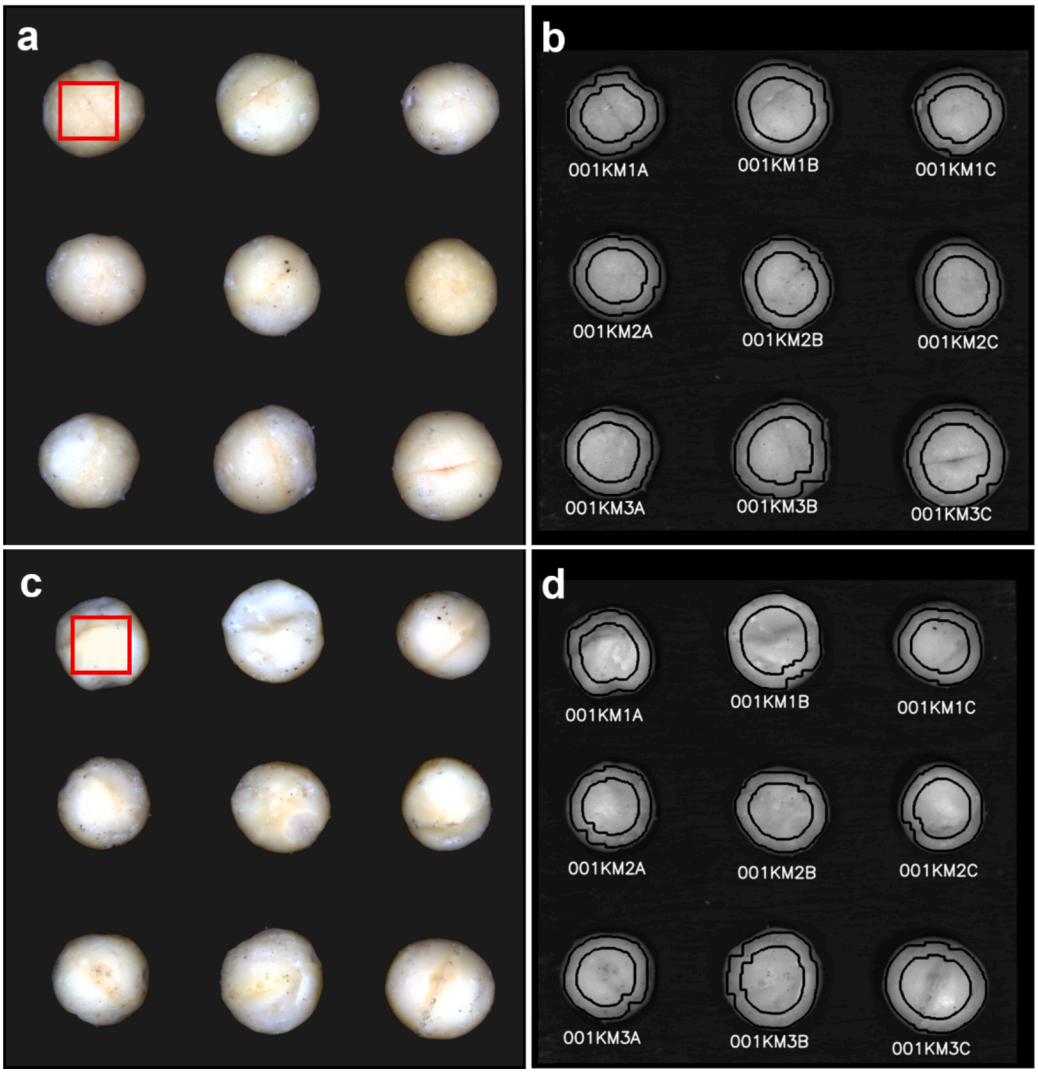


Fig. 3. (*continued*).

Thermoline, Australia). After drying for 24 hrs, samples were reweighed within 10 min of removal from the oven and moisture content MC (%) was determined by weight loss using Eq. (2):

$$MC\,\% = \left(\frac{(w - d)}{w}\right) \times 100 \quad (2)$$

where w is sample wet weight and d is sample dry weight after drying. Five nut-in-shell samples with moisture concentration greater than 25 % were manually cracked and inspected. In these cases, the kernel had turned black indicating burnt oil due to kernel immaturity or disease, These samples were discarded and removed from both the manual and the automatic datasets. Five kernel samples with moisture concentration greater than 25 % were also removed from both the manual and automatic datasets.

2.4. Spectral analysis using principal component analysis, dataset assignment and spectral pre-processing

Spectral outliers were identified and removed using principal component analysis (PCA) of all original wavelengths (n = 462) in the combined image orientation, and for nut-in-shell and kernel samples, respectively. Individual samples greater than Hotelling's T^2 0.1 % ellipse (99.9 % level of confidence) were classified as outlying and removed from both the manual and automatic datasets prior to model calibration. In total, two nut-in-shell samples were identified as spectral outliers and removed. PCA was conducted in Unscrambler software (Version 11, CAMO Software Inc., Trondheim, Norway). The remaining samples were randomly assigned (80:20) to calibration and external test datasets (Table 1). Initially PLSR models were developed to correlate moisture values with reflectance using all available untransformed wavelength bands (n = 462) from the full spectral range (400–1000 nm), and for each extraction method and image orientation, respectively (Malmir et al., 2019; Wold et al., 2001). Then a variety of ad hoc transformations were applied to spectral data to improve signal to noise ratio and prediction accuracy of the resultant models (Rinnan et al., 2009; Tahmasbian et al., 2017). The pre-processing techniques explored included standard normal variate (SNV), multiplicative scatter correction (MSC) and Savitzky-Golay (SG) first to third order derivatives with varying symmetrical smoothing point magnitudes (Fearn 2000; Kamruzzaman et al., 2016; Li et al., 2006; Rinnan et al., 2009). PLSR models were developed using *Scikit-learn* in Python 3.10 and the optimal number of latent variables were selected based on minimal mean square error (MSE) during calibration using 10-fold cross-validation (Pedregosa et al., 2011; Pelliccia, 2018).

2.5. Wavelength selection and model development

The hyperspectral images in this study contained 462 contiguous

Table 1
Descriptive statistics for macadamia moisture concentration (%) for all samples used during model development and in the training and external test datasets.

	Nut-in-shell moisture %			Kernel moisture %		
	All samples	Training dataset	External test dataset	All Samples	Training dataset	External test dataset
n	480	384	96	524	419	105
Mean	13.02	12.86	13.65	3.39	3.47	3.08
SD	5.55	5.39	6.16	4.17	4.34	3.40
Median	11.98	11.94	12.88	1.70	1.70	1.64
IQR	9.37	8.97	10.95	0.91	0.93	0.83
Skewness	0.07	0.09	−0.06	2.22	2.17	2.20
Kurtosis	−1.16	−1.09	−1.38	4.20	3.90	3.84

n: Number of samples; SD: Standard deviation; IQR: Interquartile range.

wavelengths with high correlation between adjacent bands (Fig. 4a,e). Therefore, wavelength selection was used to reduce the number of input variables, reduce redundancy, and improve model prediction accuracy (Zhu and Li, 2019). We employed two techniques to identify important wavelengths and aimed to reduce the number of wavelength bands from 462 to 10 to ensure our results could inform the development of multispectral systems and maximise the potential for industrial application in post-harvest processing (Bendel et al., 2020; Kamruzzaman et al., 2016). Initially and for models developed using PLSR only, we used a filter method to select wavelengths based on the magnitude of their respective β-coefficients calculated during initial PLSR model development and using all 462 wavelength bands (Mehmood et al., 2012; Pelliccia, 2018). The β-coefficients are the regression coefficients uniquely quantifying the influence of each wavelength band with the response variable (moisture concentration, in our case) for a given PLSR model (Kamruzzaman et al., 2016). Wavelength bands with large magnitude (positive or negative values) β-coefficients are considered important in the regression model and consequently selected for inclusion (Kamruzzaman et al., 2016). Wavelength filtering was implemented in Python by extracting regression β-coefficients from the initial PLSR model, discarding wavelengths with lowest coefficients and refitting the model (Pedregosa et al., 2011; Pelliccia, 2018). The process was repeated with the next lowest wavelength coefficient until the MSE of 10-fold cross-validation during calibration was minimised, and at this point the selection process was stopped (Pelliccia, 2018). The best-fitting models for both nut-in-shell and kernels developed in Python were then re-developed using Unscrambler software (Version 11, CAMO Software Inc., Trondheim, Norway) following selection of 10 wavelengths corresponding to the largest coefficients identified during the first step in Python. Re-developing models in Unscrambler was undertaken to validate the models developed using Python, avoid overfitting by selecting the optimal number of latent variables at the minimum predicted residual error sum of squares (PRESS), and to perform full leave-one-out cross validation by systematically removing one sample from the data set and re-fitting the model using the remaining samples (Farrar et al., 2023).

The second wavelength selection technique we used was the minimum redundancy maximum relevance (MRMR) algorithm in MATLAB (MathWorks Inc., MATLAB Version: 9.14.0.2206163 R2023a, Natick, MA) to identify important wavelengths for models developed using ANN, GPR and SVM (Peng et al., 2005; Zhu and Li, 2019). MRMR is also a type of filtering method to select wavelengths and identifies an optimal set of specified size that are mutually and maximally dissimilar, and can represent the response variable effectively (Ding and Peng, 2005; Peng et al., 2005). More specifically, MRMR quantifies the importance of a wavelength using a heuristic algorithm and returns a score. A large score value indicates that the corresponding predictor is important and wavelengths with the largest scores and both nonzero relevance and redundancy are selected up to the required number of wavelengths (MathWorks Robotics System Toolbox User's Guide, 2024). We defined the MRMR algorithm to select a subset of 10 wavelengths from the original 462 for use in subsequent model development to ensure our results were realistic for use in the development of multispectral systems (Bendel et al., 2020).

2.6. Model evaluation and selection

Best-fit calibration models were selected using the highest coefficient of determination for calibration (R^2_C) and cross-validation (R^2_{CV}) and lowest root mean square error for calibration (RMSEc) and cross-validation (RMSEcv) (Farrar et al., 2023). We compared the prediction accuracy and ratio of prediction deviation (RPD) of the best-fit calibration models developed for each fitting method and image orientation. Calibration models were used to predict moisture concentration for samples in the external test dataset and model performance was evaluated by inspecting the coefficient of determination (R^2_T) and root mean square error ($RMSE_T$) for samples in the external test dataset and ratio of prediction deviation (RPD) calculated using Eq. (3):

$$RPD = SDy/RMSEP_T, \quad (3)$$

where *SDy* is the standard deviation for moisture reference values in the external test dataset and $RMSEP_T$ is the root mean square error of prediction (Rossel and Webster, 2012). Using this metric to evaluate models developed for an RPD between 2.5 and 2.9 is 'fair' and may be used for screening applications; a value between 3.0 and 3.4 is 'good' and may be used for quality control applications; a value between 3.5 and 4.0 is 'very good' and may be used for process control applications; and a value above 4.1 is 'excellent' and suitable for any application (Williams, 2014).

3. Results

3.1. Efficiency of computer vision to extract spectral data and principal component analysis of different image orientations

Using computer vision software to automatically detect nut objects and regions of interest (ROIs) in images of nut-in-shell was possible for 97.5 % and 96.5 % of samples in image 1 and 2 orientations, respectively (Supplementary Fig. S1; Supplementary Table S1). Mean reflectance signatures for nut-in-shell imaged in the two different orientations and when combined were identical (Fig. 4a). Combining spectral data from the two image orientations resulted in marginally greater explained variance in the first principal component (91.12 %), compared with single images alone (91 % and 90.42 %, respectively) (Fig. 4b,c,d) which may be within the margins of uncertainty. Automatic extraction of spectra from images of macadamia kernels was possible for 96.8 % of samples imaged in base-up orientation and 93.9 % in base-down orientation (Fig. 2 and Supplementary Table S1). Mean spectral reflectance of kernels in base-up orientation was greater in the NIR region (800–1000 nm) and lower in the visible region (400–650 nm) compared with base-down orientation (Fig. 4e). This trend reversed for base-down orientation, and unexpectedly, this effect was averaged for both orientations combined (Fig. 4e). Spectra extracted from images of kernels in base-down orientation resulted in marginally greater explained variance

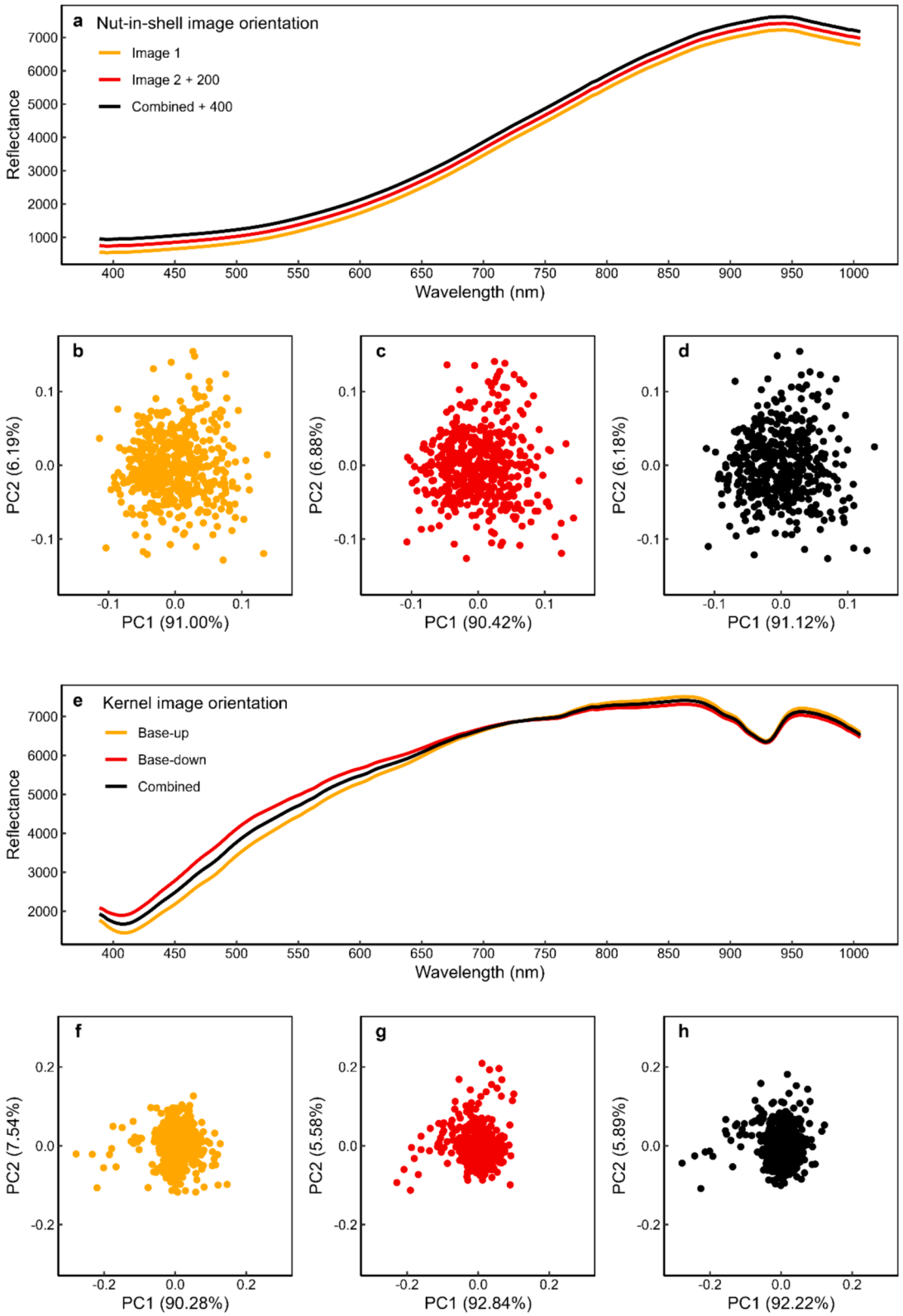


Fig. 4. **(a)** Mean reflectance collected from macadamia nut-in-shell image 1 (yellow) and image 2 (red) orientations, and image 1 and 2 combined (black) by averaging; principal component analysis scores plot for nut-in-shell in **(b)** image 1; **(c)** image 2; and **(d)** images 1 and 2 combined; **(e)** Mean reflectance collected from macadamia kernels in base-up (yellow) and base-down (red) orientations, and base-up and base-down images combined (black) by averaging; principal component analysis scores plot for kernels imaged in **(f)** base-up; **(g)** base-down; and **(h)** base-up and base-down images combined. Mean relative reflectance was scaled to 10,000 (integers) by default. Due to similarity, reflectance data in panel **(a)** for nut-in-shell image 2 and images 1 and 2 combined have been scaled by adding 200 and 400 relative reflectance integers, respectively.

in the first principal component (92.84 %), compared with base-up or combined images (90.28 % and 92.22 %, respectively) (Fig. 4f,g,h).

3.2. Prediction of moisture concentrations using spectra extracted manually and automatically with computer vision

Nut-in-shell moisture concentrations were predicted with similar accuracy using either manually or automatically extracted data and PLSR comparing all nut orientations $R^2_C > 0.90$ and $R^2_T > 0.89$, respectively (Table 2). The largest difference in prediction accuracy comparing manual to automatically extracted data and PLSR occurred when a single image orientation was used with a $\Delta R^2_T \pm 0.03$ (Table 2). Nut-in-shell moisture prediction accuracy was similar using manual or automatically extracted spectra to develop models with GPR and SVM, however, when models were developed with ANN using automatically extracted spectra increased RPD from 2.54 to 3.05 (Table 3).

Kernel moisture concentrations were predicted with similar accuracy using manually or automatically extracted spectra and PLSR comparing all nut orientations $R^2_C > 0.97$ and $R^2_T > 0.92$, respectively (Table 4). Comparing our best-fitting models developed using PLSR, automatically extracted data provided higher reliability RPD = 11.05 compared with manually extracted data RPD = 8.24 (Table 4). Kernel moisture prediction accuracy was similar for models developed using manual or automatically extracted spectra with ANN, GPR and SVM (Table 5). However, RPD increased from > 6 to > 8 for models developed using automatically extracted spectra in combination with ANN and GPR (Table 5).

3.3. Prediction of moisture concentrations using spectra extracted from different nut orientations and both images combined

Nut-in-shell moisture concentrations were predicted with $R^2_T > 0.90$ using data extracted from either nut orientation (Table 2). Combining the data from two images improved nut-in-shell prediction accuracy and reliability compared with using data from a single image only. For example, models developed using all transformed spectra combined from the two different images with PLSR increased R^2_T from 0.88 to 0.92 and RPD from 2.9 to 3.5 compared with using data from a single image only (Table 2).

Kernel moisture concentrations were predicted with $R^2_T > 0.92$ and RPD > 3.5 using data extracted from images of kernels in either base-up or base-down orientations or when combining data from both orientations with PLSR (Table 4). However, models developed using combined data from both orientations had higher prediction accuracy $R^2_T = 0.99$ and RPD > 11 compared with either base-up or base-down image orientations (Table 4).

3.4. Important and overlapping wavelengths corresponding to prediction of macadamia moisture concentration

Spectral regions with highest absolute weighted β-coefficients for prediction of nut-in-shell moisture were located at approximately 400–420, 450–500, 520–570, 750–775, 815, 850–920 and 950–990 nm (Supplementary Table S2). However, the most parsimonious model was developed using 10 wavelength bands located at approximately 760, 880, and between 950 and 970 nm (Fig. 5; Supplementary Table S2). The PLSR model re-developed using only 10 wavelengths transformed with Savitsky-Golay 2nd derivative reduced the number of required latent variables to 3 (from 12 that were required using the entire wavelength range) while maintaining calibration accuracy $R^2_C = 0.87$, prediction accuracy $R^2_T = 0.88$, $RMSE_T = 2.14$ %, and test reliability RPD = 2.88 (Fig. 5).

Spectral regions with highest absolute weighted β-coefficients for prediction of kernel moisture were located at approximately 400, 450–570, 650, 750–770, 780, 860–870 and 900–1000 nm (Supplementary Table S2). However, the most parsimonious model was developed using 10 wavelengths located at approximately 780 and between 950 and 970 nm (Fig. 6; Supplementary Table S2). The best PLSR model re-developed using only 10 wavelengths transformed with Savitsky-Golay 2nd derivative reduced the number of required latent variables

Table 2
Performance of PLSR models developed using two different Vis/NIR hyperspectral image orientations and combined images of macadamia nut-in-shell to predict moisture concentration (%) using raw spectral data, data transformed with Savitsky-Golay smoothing and following selection of important wavelengths.

Orientation	Transformation	WLS	WL	LV	R^2_C	R^2_{cv}	$RMSE_C$	$RMSE_{CV}$	R^2_T	$RMSE_T$	RPD
Manual extraction of spectral data											
First image	Raw spectra	All	462	16	0.90	0.87	1.67	1.92	0.89	2.00	3.08
		Selected	172	15	0.90	0.88	1.67	1.87	0.95	1.34	4.59
	SG $^{2nd}{}_{17pt}$	All	462	11	0.91	0.88	1.62	1.91	0.91	1.81	3.41
		Selected	118	11	0.91	0.89	1.58	1.79	0.96	1.22	5.04
Rotated image	Raw spectra	All	462	15	0.90	0.87	1.69	1.93	0.90	1.95	3.17
		Selected	125	15	0.90	0.88	1.69	1.87	0.95	1.41	4.37
	SG $^{2nd}{}_{17pt}$	All	462	8	0.90	0.87	1.72	1.92	0.90	1.91	3.23
		Selected	146	8	0.90	0.88	1.70	1.90	0.95	1.37	4.49
Combined images	Raw spectra	All	462	17	0.91	0.88	1.61	1.85	0.92	1.77	3.49
		Selected	106	15	0.91	0.89	1.64	1.80	0.95	1.32	4.67
	SG $^{2nd}{}_{17pt}$	All	462	12	0.92	0.88	1.57	1.84	0.91	1.79	3.44
		Selected	90	12	0.92	0.90	1.54	1.73	0.96	1.20	5.15
Automatic extraction of spectral data											
First image	Raw spectra	All	462	15	0.90	0.88	1.665	1.888	0.89	1.93	3.19
		Selected	102	15	0.91	0.88	1.642	1.849	0.95	1.37	4.50
	SG $^{2nd}{}_{17pt}$	All	462	11	0.91	0.88	1.582	1.853	0.88	2.11	2.93
		Selected	97	11	0.92	0.90	1.553	1.742	0.95	1.36	4.54
Rotated image	Raw spectra	All	462	16	0.91	0.88	1.629	1.876	0.90	1.91	3.23
		Selected	263	16	0.91	0.89	1.607	1.828	0.96	1.19	5.16
	SG $^{2nd}{}_{17pt}$	All	462	9	0.91	0.88	1.650	1.871	0.88	2.15	2.86
		Selected	153	9	0.91	0.89	1.579	1.811	0.96	1.30	4.75
Combined images	Raw spectra	All	462	16	0.91	0.89	1.583	1.801	0.92	1.74	3.53
		Selected	120	15	0.91	0.89	1.580	1.758	0.96	1.21	5.10
	SG $^{2nd}{}_{17pt}$	All	462	12	0.92	0.89	1.541	1.788	0.92	1.75	3.53
		Selected	85	12	0.92	0.90	1.513	1.696	0.96	1.21	5.11

WL: Number of wavelengths used in model calibration; LV: Optimal number of latent variables used in the model; SG: Savitzky-Golay derivative transformation with polynomial order and smoothing range described by super and subscript respectively; OSC: Orthogonal signal correction; RPD: Ratio of prediction to deviation.

Table 3
Comparison of machine learning models developed using data manually and automatically extracted and combined from two separate hyperspectral images to predict macadamia nut-in-shell moisture concentration (%) using all wavelengths (n = 462) and following selection of ten important wavelengths using MRMR and β-coefficient filtering methods.

Model	Details	WL	R^2_{cv}	$RMSE_{CV}$	R^2_T	$RMSE_T$	RPD
Manual extraction of spectral data							
ANN	Narrow ANN (1 layer, 10 nodes, ReLU activation, λ = 0)	462	0.86	2.13	0.84	2.07	2.54
	MRMR	10	0.64	3.67	0.54	3.56	1.48
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.73	2.95	0.64	3.16	1.67
GPR	Matérn 5/2 (isotropic kernel)	462	0.90	1.82	0.90	1.66	3.18
	MRMR	10	0.69	3.14	0.70	2.89	1.83
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.73	2.92	0.68	2.98	1.77
SVM	Quadratic	462	0.76	2.75	0.77	2.51	2.10
	MRMR	10	0.54	3.81	0.52	3.61	1.46
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.66	3.29	0.58	3.41	1.55
Automatic extraction of spectral data							
ANN	Narrow ANN (1 layer, 10 nodes, ReLU activation, λ = 0)	462	0.88	1.92	0.89	1.82	3.05
	MRMR	10	0.73	2.92	0.73	2.86	1.94
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.72	2.96	0.72	2.90	1.92
GPR	Matérn 5/2 (isotropic)	462	0.90	1.79	0.91	1.63	3.41
	MRMR	10	0.75	2.80	0.71	2.96	1.88
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.73	2.88	0.74	2.80	1.99
SVM	Quadratic	462	0.77	2.67	0.78	2.58	2.15
	MRMR	10	0.65	3.29	0.49	3.96	1.40
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.65	3.32	0.69	3.09	1.80

ANN: Artificial neural network; GPR: Gaussian process regression; MRMR: Minimum redundancy maximum relevance; ReLU: Rectified linear unit; SVM: Support vector machine; WL: Wavelengths.

to 1 (from 18) while maintaining calibration accuracy $R^2_C = 0.93$, prediction accuracy $R^2_T = 0.92$, $RMSE_T = 0.92$ %, and test reliability RPD = 4.26 (Fig. 6). All wavelengths selected during model development and corresponding to prediction of moisture for nut-in-shell and kernel samples are presented in (Supplementary Table S2).

3.5. Comparison of machine learning models to predict macadamia nut-in-shell and kernel moisture concentration

The best model we developed to predict nut-in-shell moisture concentration was developed using PLSR with data extracted from both images combined and following selection of 90 wavelengths ($R^2_T = 0.96$, $RMSE_T = 1.21$ %, and RPD = 5.11) (Table 2). Models developed using ANN and GPR and all wavelengths also resulted in high accuracy ($R^2_T > 0.85$, $RMSE_T < 2.1$ %) moisture prediction using data either manually or automatically extracted (Table 3). However, re-developing the ANN, GPR and SVM models using ten wavelengths selected by MRMR, and again using the same ten wavelengths identified during PLSR (Supplementary Table S2), prediction accuracy was reduced $R^2_T \leq 0.75$, $RMSE_T \geq 2.80$ % and RPD < 2 (Table 3). The best model calibrated using SVM predicted moisture with $R^2_T = 0.78$, $RMSE_T = 2.58$ % and required all available wavelengths (Table 3).

The best model to predict kernel moisture was developed using PLSR and with data extracted from both images combined and following selection of 140 wavelengths ($R^2_T = 0.99$, $RMSE_T = 0.31$ % and RPD = 11.05) (Table 4). Models developed using ANN, GPR and SVM and all wavelengths also resulted in high accuracy ($R^2_T \geq 0.96$, $RMSE_T \leq 1.03$ %) prediction of kernel moisture (Table 5). However, re-developing the ANN, GPR and SVM models using ten wavelengths selected by MRMR, and again using the same ten wavelengths identified during PLSR (Supplementary Table S2), did not significantly reduce prediction accuracy $R^2_T > 0.79$, $RMSE_T \leq 2.09$ % and RPD > 2.2 (Table 5). Additionally, the ANN model developed using the most important wavelengths identified during PLSR by combining β-coefficient filtering and selection of wavelengths with large magnitude were comparable to the model developed using all available wavelengths (Table 5).

3.6. Important and overlapping wavelengths corresponding to prediction of macadamia moisture concentration

Spectral regions with highest absolute weighted β-coefficients for prediction of nut-in-shell moisture were located at approximately 400–420, 450–500, 520–570, 750–775, 815, 850–920 and 950–990 nm (Supplementary Table S2). However, the most parsimonious model was developed using 10 wavelength bands located at approximately 760, 880, and between 950 and 970 nm (Fig. 5; Supplementary Table S2). The PLSR model re-developed using only 10 wavelengths transformed with Savitsky-Golay 2nd derivative reduced the number of required latent variables to 3 (from 12 that were required using the entire wavelength range) while maintaining calibration accuracy $R^2_C = 0.87$, prediction accuracy $R^2_T = 0.88$, $RMSE_T = 2.14$ %, and test reliability RPD = 2.88 (Fig. 5).

Spectral regions with highest absolute weighted β-coefficients for prediction of kernel moisture were located at approximately 400, 450–570, 650, 750–770, 780, 860–870 and 900–1000 nm (Supplementary Table S2). However, the most parsimonious model was developed using 10 wavelengths located at approximately 780 and between 950 and 970 nm (Fig. 6; Supplementary Table S2). The best PLSR model re-developed using only 10 wavelengths transformed with Savitsky-Golay 2nd derivative reduced the number of required latent variables to 1 (from 18) while maintaining calibration accuracy $R^2_C = 0.93$, prediction accuracy $R^2_T = 0.92$, $RMSE_T = 0.92$ %, and test reliability RPD = 4.26 (Fig. 6). All wavelengths selected during model development and corresponding to prediction of moisture for nut-in-shell and kernel samples are presented in (Supplementary Table S1).

4. Discussion

4.1. Moisture prediction for macadamia nut-in-shell and kernels

Hyperspectral imaging successfully predicted moisture concentration in both macadamia nut-in-shell and kernel samples. Macadamia fruit consist of an outer husk that encloses the nut-in-shell with a kernel located inside (Trueman, 2013). We predicted both nut-in-shell and kernel moisture concentration with high prediction accuracy. In general, selecting important wavelengths improved model accuracy and the

Table 4
Performance of PLSR models developed using two different Vis/NIR hyperspectral image orientations and combined images of macadamia kernels to predict moisture (%) concentration using raw spectral data, data transformed with Savitsky-Golay smoothing and following selection of important wavelengths.

Orientation	Transformation	WLS	WL	LV	R^2_C	R^2_{cv}	$RMSE_C$	$RMSE_{CV}$	R^2_T	$RMSE_T$	RPD
Manual extraction of spectral data											
Base-up	Raw spectra	All	462	20	0.98	0.97	0.585	0.723	0.97	0.629	5.41
		Selected	138	14	0.98	0.97	0.641	0.711	0.98	0.471	7.22
	$SG^{2nd}{}_{17pt}$	All	462	13	0.99	0.97	0.536	0.784	0.95	0.731	4.66
		Selected	139	13	0.98	0.97	0.578	0.704	0.99	0.341	9.98
Base-down	Raw spectra	All	462	16	0.97	0.96	0.799	0.902	0.92	0.958	3.55
		Selected	141	14	0.97	0.96	0.793	0.870	0.97	0.575	5.92
	$SG^{2nd}{}_{17pt}$	All	462	12	0.97	0.96	0.737	0.899	0.93	0.881	3.86
		Selected	101	11	0.97	0.96	0.740	0.841	0.98	0.494	6.89
Combined	Raw spectra	All	462	19	0.98	0.97	0.584	0.702	0.96	0.690	4.93
		Selected	32	19	0.98	0.98	0.605	0.673	0.98	0.452	7.53
	$SG^{2nd}{}_{17pt}$	All	462	14	0.98	0.97	0.554	0.704	0.96	0.696	4.89
		Selected	65	13	0.98	0.98	0.575	0.650	0.99	0.413	8.24
Automatic extraction of spectral data											
Base-up	Raw spectra	All	462	19	0.98	0.97	0.600	0.747	0.96	0.647	5.26
		Selected	115	19	0.98	0.97	0.564	0.700	0.99	0.366	9.30
	$SG^{2nd}{}_{17pt}$	All	462	15	0.98	0.97	0.567	0.787	0.97	0.610	5.58
		Selected	129	14	0.98	0.97	0.565	0.696	0.99	0.341	9.98
Base-down	Raw spectra	All	462	19	0.97	0.96	0.776	0.916	0.93	0.930	3.66
		Selected	216	18	0.97	0.96	0.772	0.883	0.98	0.496	6.86
	$SG^{2nd}{}_{17pt}$	All	462	12	0.97	0.96	0.745	0.916	0.93	0.902	3.77
		Selected	135	7	0.97	0.96	0.779	0.877	0.97	0.577	5.90
Combined	Raw spectra	All	462	20	0.98	0.97	0.581	0.701	0.96	0.706	4.82
		Selected	113	20	0.98	0.98	0.556	0.666	0.99	0.354	9.61
	$SG^{2nd}{}_{17pt}$	All	462	18	0.99	0.97	0.524	0.710	0.95	0.757	4.50
		Selected	140	18	0.99	0.98	0.522	0.641	0.99	0.308	11.05

WL: Number of wavelengths used in model calibration; LV: Optimal number of latent variables used in the model; SG: Savitzky-Golay derivative transformation with polynomial order and smoothing range described by super and subscript respectively; OSC: Orthogonal signal correction; RPD: Ratio of prediction to deviation.

models developed using only 10 wavelength bands had marginally lower prediction accuracy but were still acceptable for use in 'screening applications' of nut-in-shell and 'process control applications' for macadamia kernels (Williams, 2014).

Macadamia moisture concentration was predicted with high accuracy using only ten selected wavelength bands. Successful prediction of moisture can be explained by the important wavelengths that were identified in the region related to NIR absorption due to O–H bonds present in water molecules. In this study, the most important wavelengths used to predict macadamia moisture were located between 950 and 970 nm and at 760 nm for nut-in-shell and 780 for kernels. Wavelengths in the 950–970 nm region are associated with water absorption and have been widely utilised in the prediction of water content and moisture in various soils and crops (ElMasry et al., 2007; Huang et al., 2014; Wei et al., 2019; Wu et al., 2012; Yi et al., 2013). Bending/stretch vibration of hydroxy (O–H) bonds in water molecules has been related to absorption bands at 970 nm in shelled almond kernels (Panda et al., 2022), 940 nm in soil (Qi et al., 2017), and in fruit between 930 and 1080 nm (Qing et al., 2007). Furthermore, wavelengths at 958 nm and 760 nm were identified as two of 12 optimum wavelengths selected to predict moisture content of prawns (Wu et al., 2012). Our results are also consistent with high accuracy prediction of moisture concentrations in almonds and soybeans using PLSR and mean reflectance of selected wavelengths around 960–970 nm (Huang et al., 2014, Panda et al., 2022). Future development of hyperspectral to predict macadamia moisture method should exploit the wavelengths we identified in this study.

4.2. Efficacy of computer vision to automatically detect macadamia nuts and extract spectral data

Using a computer vision method based on contour detection led to successful automatic feature extraction of macadamia nut-in-shell and kernels from within hyperspectral images, and the models we developed using automatically extracted spectra generally predicted nut moisture with comparable accuracy to using data extracted manually. Manual feature extraction is the process of selecting features using an images' information and involved the user manually selecting (handcrafted) a rectangular/square ROI shape within the kernel boundary of the digital image. Conversely, the automatic method identified circular/elliptical shapes between the nuts and the image background based on captured local contrasts by applying contour detection (Ma et al., 2021). Both extraction methods led to comparable moisture prediction for macadamia nut-in-shell because they are generally close to spherical. We suggest that using spectral data automatically extracted in a circle/ellipse shape is more reliable for nut moisture prediction than using data from rectangle/square ROIs manually selected by a human operator. We also found that for kernels with a morphologically flatter base and pointed crest/shoulder, the automatic method led to higher prediction accuracy when the extracted spectral data from both images were combined by averaging. This indicates that adopting automatic computer vision systems to extract spectral data during macadamia processing will not limit model prediction accuracy. Further improvement in model accuracy may be possible by implementing more sophisticated spectral-spatial boundary detection algorithms rather than using contour detection during data extraction (Al-Khafaji et al., 2021).

4.3. The effect of nut orientation during imaging on moisture prediction

Moisture prediction accuracy for nut-in-shell was similar using the spectral data extracted from a single nut image or two images combined, however for kernels, prediction accuracy was higher using images of base-up kernels compared with base-down. Macadamia nut-in-shells are round, and kernels are also generally round, however, one side of each kernel has a defined and pointed crest/shoulder that develops at the micropylar surface, and the opposite basal side has a flatter ovoid shape (Trueman, 2013; Walton and Wallace, 2008). Therefore, similar prediction accuracy for the two nut-in-shell images was expected given the regular rounded shape and similar reflectance scattering for both orientations. Whereas higher prediction accuracy using images of kernels

Table 5
Comparison of machine learning models developed using data manually and automatically extracted and combined from two separate hyperspectral images to predict macadamia kernel moisture concentration (%) using all wavelengths (n = 462) and following selection of ten important wavelengths using MRMR and β-coefficient filtering methods.

Model	Details (layers, size, activation, lambda)	WL	R^2_{cv}	$RMSE_{CV}$	R^2_T	$RMSE_T$	RPD
Manual extraction of spectral data							
ANN	Bi-layered (2, 10, 10, ReLU, 0)	462	0.96	0.82	0.97	0.73	6.28
	MRMR	10	0.77	1.96	0.89	1.50	3.03
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.98	0.64	0.97	0.78	5.87
GPR	Matérn 5/2 (isotropic)	462	0.98	0.56	0.97	0.74	6.19
	MRMR	10	0.84	1.61	0.92	1.26	3.63
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.92	1.17	0.94	1.08	4.24
SVM	Quadratic	462	0.95	0.91	0.97	0.74	6.18
	MRMR	10	0.64	2.45	0.79	2.09	2.19
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.91	1.21	0.95	0.97	4.69
Automatic extraction of spectral data							
ANN	Narrow (1, 10, ReLU, 0)	462	0.97	0.71	0.99	0.56	8.81
	MRMR	10	0.90	1.33	0.93	1.14	3.78
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.95	0.85	0.98	0.72	6.81
GPR	Squared exponential (isotropic kernel)	462	0.97	0.69	0.99	0.57	8.69
	MRMR	10	0.51	2.90	0.91	1.26	3.41
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.85	1.52	0.95	1.12	4.41
SVM	Quadratic	462	0.95	0.91	0.96	1.03	4.79
	MRMR	10	0.94	0.99	0.95	0.93	4.64
	β-coefficient filtering during PLSR + select 10 with largest magnitude	10	0.93	1.08	0.95	1.12	4.39

ANN: Artificial neural network; GPR: Gaussian process regression; MRMR: Minimum redundancy maximum relevance; ReLU: Rectified linear unit; SVM: Support vector machine; WL: Wavelengths.

in base-up orientation is most likely due to the flatter and more regular base-up surface resulting in a similar optical path compared with a pointed crest scattering light away from the hyperspectral sensor. Hyperspectral imaging has previously been developed to assess the moisture concentration and distribution in several other foods including almond nuts (Panda et al., 2022), prawns (Wu et al., 2012), red meat (Kamruzzaman et al., 2012), nuts and grains (Mohammadi-Moghaddam et al., 2018) whose shapes are also irregular or spherical. Therefore, our results are consistent with reliable moisture prediction in food products.

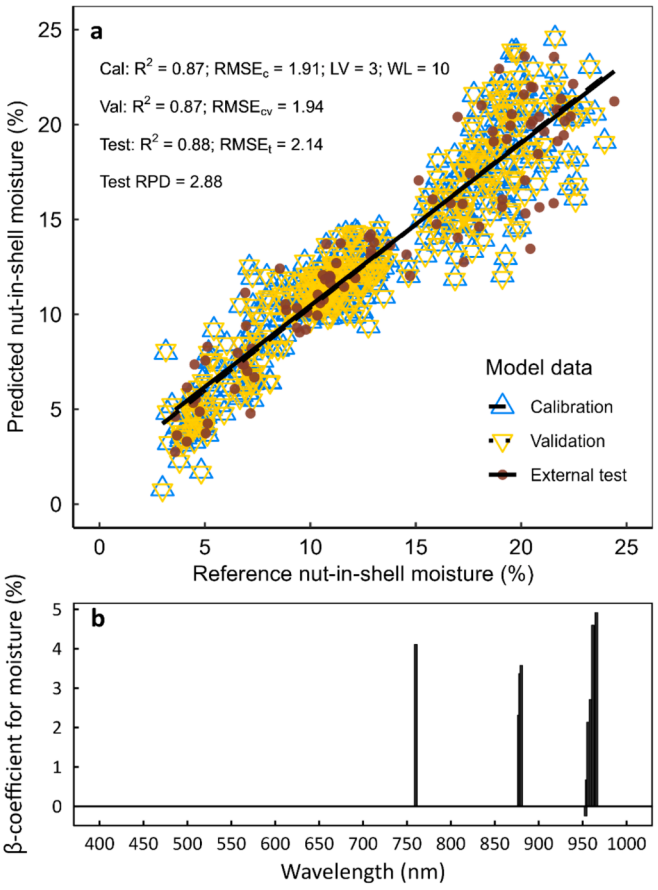


Fig. 5. **(a)** Best-fitting PLSR model developed using images of macadamia nut-in-shell following selection of 10 wavelengths and showing measured *vs.* predicted moisture (%) concentration in the calibration data (Calibration): upright blue triangle; validation data (Validation): upside-down yellow triangle; and external test data sets: closed brown circle (External test) respectively; and **(b)** weighted β-coefficients for wavelengths selected from all 462 wavelengths available and prior to development of best-fitting PLSR model depicted in (a).

4.4. Comparison of machine learning models and important wavelength bands for predicting macadamia moisture concentration

In general, PLSR provided higher or comparable moisture prediction accuracy than the respective ANN. However, ANN developed using the ten important wavelengths identified during PLSR development provided higher prediction accuracy than the PLSR counterpart for kernel samples only. This was unexpected given that the method used to select wavelengths during PLSR calibration was based on filtering β-coefficients and decomposition of spectra into latent variables, and is native to PLSR not ANN. However, the PLSR method and decomposition of spectra into latent variables is well suited to reduce collinearity and is robust against a reduction in wavelengths, and therefore, may be useful to identify important wavelengths prior to development models that so not utilise latent variables or β-coefficients natively such as ANN.

Our best-fit PLSR models provided R^2 values between 0.87 and 0.99, for predicting nut-in-shell and kernel moisture concentrations. R^2 values between 0.83 and 0.98, are potentially useful for quality assurance and research purpose applications with caution (Williams et al., 2019). PLSR and ANN are two commonly used algorithms for modelling hyperspectral data. PLSR is especially useful when spectral data and chemical reference values are related in a linear fashion and when the size of image datasets are small (Rossel, 2008). On the other hand, ANNs can handle non-linear relations between spectra and reference values, but require larger image datasets for robust performance, and decreasing

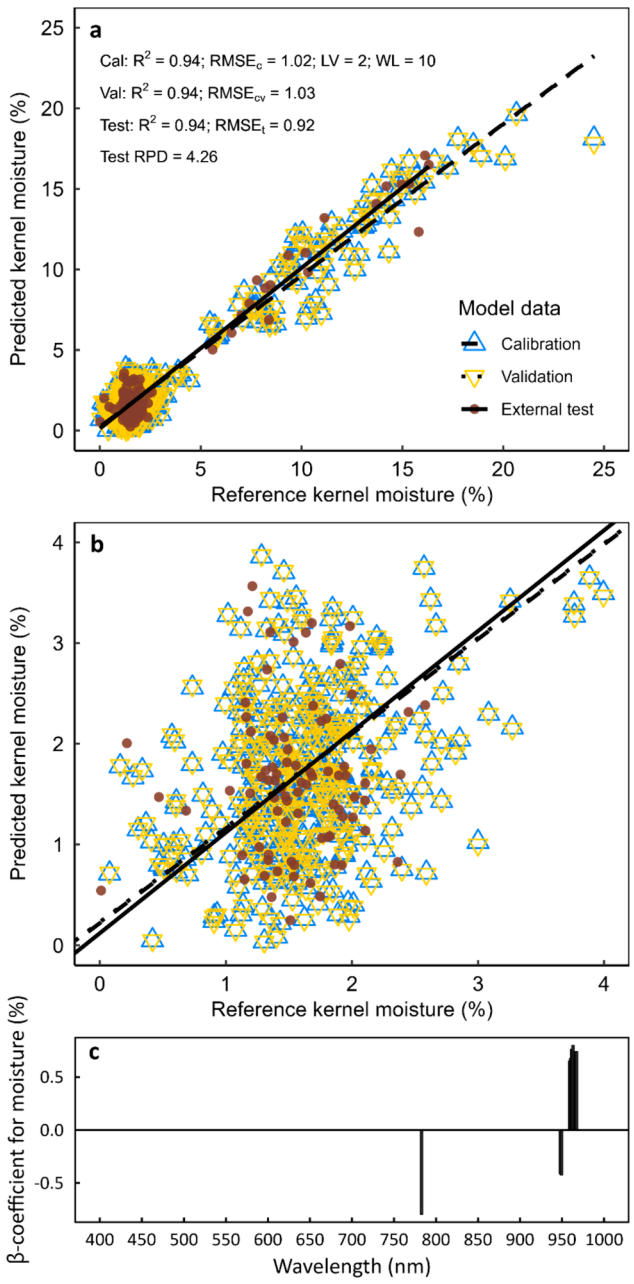


Fig. 6. Best-fitting PLSR model developed using images of macadamia kernels following selection of 10 wavelengths and showing measured *vs.* predicted moisture (%) concentration in the calibration data (Calibration): upright blue triangle; validation data (Validation): upside-down yellow triangle; and external test data sets: closed brown circle (External test) respectively, **(a)** showing the entire data range between 0 and 25% moisture; **(b)** showing a reduced and scaled range between 0 and 4% moisture only; and **(c)** weighted β-coefficients for wavelengths selected from all 462 wavelengths available and prior to development of best-fitting PLSR model depicted in (a).

dataset size can also reduce model prediction reliability (Bian et al., 2012; Habibi et al., 2020). In our study, PLSR provided higher moisture prediction accuracy than ANN models suggesting that a linear relation between water absorption features of the spectra and moisture concentration does exist in our current dataset.

4.5. Implications and limitation of this scoping study and future recommendations

Prediction accuracy for models developed with data automatically extracted from hyperspectral images using computer vision were similar compared to models developed with manually extracted data. Automated sorting and grading systems that operate using RGB bands only are already widely used in food and nut processing industries around the world to improve processing efficiency. Colour sorting machines can identify physical defects such as mould, insect damage and to separate shell from kernel following cracking (MM, 2024). Therefore, the successful development of hyperspectral imaging systems that can assess other quality parameters, such as moisture concentration, in addition to physical defects would represent a significant advancement and a sensible option to retrofit to existing infrastructure. One of the challenges of image processing is to fast-track data processing and a critical step during model calibration is to identify the most important wavelengths for predicting the optical property of interest (Adebayo et al., 2016). Therefore, this study determined the optimum 10 wavelengths to remove spectral redundancy, reduce model complexity and in-turn demand on computational resources (ElMasry et al., 2012; Kamruzzaman et al., 2013). By successfully identifying 10 wavelength bands only, our results inform future development of simpler and more cost-effective multispectral systems and fast-track data processing when the technology is developed for conveyor belts in nut processing factories (Bandara et al., 2020; ElMasry et al., 2012; Mohammadi-Moghaddam et al., 2018). Additionally, future studies should also investigate using models to predict individual pixels within images and thereby mapping the intra-kernel moisture distribution allowing for identification of correlations between moisture distribution and initiation of other quality defects such as oxidative rancidity and/or brown centre disorder (Panda et al., 2022).

Currently, macadamia moisture is assessed using two methods: 1) the thermogravimetric (oven-drying) method for nut-in-shell, and 2) purpose made moisture meters that are based in a laboratory for analysis of kernels (AMS, 2021; Guthrie et al., 2004). Neither of these methods can be used on a sorting line. The oven-drying method involves drying a representative sample of nut-in-shell or kernel for 24 hrs at 105 °C and calculating loss of weight (water) and benchtop moisture meters analyse a 1 kg representative sample of macadamia nut-in-shell or kernel and report moisture values between 2–10 % and 1–8 %, respectively (AMS, 2022; GmbH, 2023). A new method that can accurately determine moisture for each individual nut, and in real-time, would represent a significant step forward for the precision of macadamia post-harvest processing. In this study, the sampling strategy was designed to capture as much variation in macadamia nut samples as possible because moisture is a critical parameter during all post-harvest processing stages. Freshly harvested nuts can have moisture concentrations $>$ 20 % and may range up to 30 % when wet (Mason, 2000; O'Hare et al., 2004). Furthermore, nuts are partially dried on-farm and delivered to processors at various moisture concentrations, therefore, it is critical that a representative sample of delivered nuts are assessed for moisture to inform grower payment calculations that are adjusted for 10 % moisture (Walton and Wallace, 2011). This study indicates that nut-in-shell moisture assessment in real-time on receipt of delivery to a processor may be possible, and therefore, future adoption will help inform process control decisions such as consolidating deliveries based on moisture concentration and grower payments. However, several physical conditions were identified to be problematic for automatic detection using machine vision (Supplementary Fig. S1). Accordingly, any real-world application of the automatic detection method would result in nuts passing through non-detected or detected poorly and continuing down a processing line without a reliable moisture concentration assessment. Therefore, a need may arise to identify and filter any problematic samples prior to moisture determination for manual assessment depending on the point during processing and the importance for an accurate moisture assessment.

5. Conclusion

Hyperspectral imaging combined with regression methods predicted the moisture concentration of nut-in-shell and kernels. This study demonstrates that reflectance values can be automatically extracted from hyperspectral images using computer vision and utilised to develop regression models to predict macadamia nut moisture concentration. Using data extracted from two images combined provided more reliable results, however, one image only of nut-in-shell may also be used, which is key for in-line sorting applications. Our results indicate that data from images of nuts in mixed orientations can still provide high prediction accuracy, and therefore, this technology can be applied to mechanised processing where nuts are transited on conveyor belts. A PLSR model developed using only 10 wavelength bands provided prediction accuracy levels that were usable with caution while reducing model complexity and computational demand. This research indicates that hyperspectral reflectance Vis/NIR spectral range can be used to quantify moisture concentration of macadamia at various points during post-harvest processing.

CRediT authorship contribution statement

Michael B. Farrar: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Reza Omidvar:** Writing – review & editing, Software, Investigation. **Joel Nichols:** Writing – review & editing, Resources, Project administration. **Daniele Pelliccia:** Writing – review & editing, Software, Formal analysis. **Suhad Lateef Al-Khafaji:** Writing – review & editing, Validation, Software, Conceptualization. **Iman Tahmasbian:** . **Nimanie Hapuarachchi:** Writing – review & editing. **Shahla Hosseini Bai:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Shahla Hosseini Bai reports financial support was provided by The Queensland Government, Department of Tourism, Innovation and Sport. Shahla Hosseini Bai reports financial support was provided by CL Macs. Shahla Hosseini Bai reports financial support was provided by Macdamias Direct. Shahla Hosseini Bai reports financial support was provided by Marquis Macdamias. Shahla Hosseini Bai reports financial support was provided by Suncoast Gold Macadamias. Shahla Hosseini Bai reports a relationship with Australian Macadamia Society that includes: non-financial support. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

Financial support for this research was provided by: The Queensland Government, Department of Tourism, Innovation and Sport (Advance QLD Industry Fellowship No: AQIRF092-2021RD4), Australian Macadamia Society, CL Macs, Cropwatch Independent Laboratory, Macadamias Direct, Marquis Macadamias, Suncoast Gold Macadamias. The authors would like to express their sincere gratitude to Helen Wallace, Stephen Trueman, Steve Lee, Joel Michael, Peter Zummo, Jeff Clements, Mariana Proksch, Jon Perrin, Michael Green, Lucy Andrews, Julian Lancaster-Smith, Royce Alcorn for sharing expert knowledge, help in guidance of the project and technical support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2024.109209.

References

Adebayo, S.E., Hashim, N., Abdan, K., Hanafi, M., 2016. Application and potential of backscattering imaging techniques in agricultural and food processing - A review. J. Food Eng. 169, 155–164.
Al-Khafaji, S.L., Zhou, J., Bai, X., Qian, Y., Liew, A.-W.-C., 2021. Spectral-spatial boundary detection in hyperspectral images. IEEE Trans. Image Processing 31, 499–512.
Ams, 2021. Kernel Assessment Manual. Australian Macadamia Society, Lismore, Australia.
Ams, 2022. Kernel Assessment Workshops, Adapted from the AMS - Scheme Rules Version 2. Australian Macadamia Society, Lismore, Australia, Kernel Assessment Workshops.
Bandara, W.G.C., Prabhath, G.W.K., Dissanayake, D.W.S.C.B., Herath, V.R., Godaliyadda, G.M.R.I., Ekanayake, M.P.B., Demini, D., Madhujith, T., 2020. Validation of multispectral imaging for the detection of selected adulterants in turmeric samples. J. Food Eng. 266, 109700.
Bendel, N., Backhaus, A., Kicherer, A., Köckerling, J., Maixner, M., Jarausch, B., Biancu, S., Klück, H.-C., Seiffert, U., Voegele, R.T., 2020. Detection of two different grapevine yellows in Vitis vinifera using hyperspectral imaging. Remote Sens. 12, 4151.
Bradski, G., 2000. The openCV library, Dr. Dobb's Journal: Software Tools for the Professional Programmer. 25, 120–123.
Buthelezi, N.M.D., Tesfay, S.Z., Ncama, K., Magwaza, L.S., 2019. Destructive and non-destructive techniques used for quality evaluation of nuts: A review. Sci. Hortic. 247, 138–146.
Ding, C., Peng, H., 2005. Minimum redundancy feature selection from microarray gene expression data. J. Bioinf. Comput. Biol. 3, 185–205.
Dung, C.D., Trueman, S.J., Wallace, H.M., Farrar, M.B., Gama, T., Tahmasbian, I., Bai, S. H., 2023. Hyperspectral imaging for estimating leaf, flower, and fruit macronutrient concentrations and predicting strawberry yields. Environ. Sci. Pollut. Res. 30, 114166–114182.
ElMasry, G., Kamruzzaman, M., Sun, D.W., Allen, P., 2012. Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: A review. Crit. Rev. Food Sci. Nutr. 52, 999–1023.
ElMasry, G., Sun, D.W., Allen, P., 2013. Chemical-free assessment and mapping of major constituents in beef using hyperspectral imaging. J. Food Eng. 117, 235–246.
Farrar, M.B., Wallace, H.M., Brooks, P., Yule, C.M., Tahmasbian, I., Dunn, P.K., Hosseini Bai, S., 2021. A performance evaluation of Vis/NIR hyperspectral imaging to predict curcumin concentration in fresh turmeric rhizomes. Remote Sens. 13, 1807.
Farrar, M.B., Wallace, H.M., Tahmasbian, I., Yule, C.M., Dunn, P.K., Hosseini Bai, S., 2023. Rapid assessment of soil carbon and nutrients following application of organic amendments. Catena 223, 106928.
Gama, T., Wallace, H.M., Trueman, S.J., Hosseini-Bai, S., 2018. Quality and shelf life of tree nuts: A review. Sci. Hortic. 242, 116–126.
GmbH, S.M., 2023. Humimeter FSG Nut Moisture Meter. Schaller Messtechnik GmbH, Ruprecht an der Raab, Austria. Available online at: < https://www.humimeter.com/en/foods/humimeter-fsg/> (Accessed 8 March 2024).
Guthrie, J., Greensill, C., Bowden, R., Walsh, K., 2004. Assessment of quality defects in macadamia kernels using NIR spectroscopy. Aust. J. Agric. Res. 55, 471–476.
Hapuarachchi, N.S., Trueman, S.J., Kämper, W., Farrar, M.B., Wallace, H.M., Nichols, J., Bai, S.H., 2023. Hyperspectral imaging of adaxial and abaxial leaf surfaces for rapid assessment of foliar nutrient concentrations in Hass avocado. Remote Sens. 15, 3100.
Huang, M., Wang, Q., Zhang, M., Zhu, Q., 2014. Prediction of color and moisture content for vegetable soybean during drying using hyperspectral imaging technology. J. Food Eng. 128, 24–30.
INC, 2023. Nuts & Dried Fruits Statistical Yearbook 2022/2023. International Nut & Dried Fruit, Reus, Spain. Available online at: <https://www.nutfruit.org/industry/technical-resources?category=statistical-yearbooks> (Accessed 8 March 2024).
Kämper, W., Trueman, S.J., Tahmasbian, I., Bai, S.H., 2020. Rapid determination of nutrient concentrations in Hass avocado fruit by Vis/NIR hyperspectral imaging of flesh or skin. Remote Sens. 12, 3409.
Kamruzzaman, M., Elmasry, G., Sun, D.W., Allen, P., 2012. Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. Innov. Food Sci. Emerg. Technol. 16, 218–226.
Kamruzzaman, M., ElMasry, G., Sun, D.W., Allen, P., 2013. Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. Food Chem. 141, 389–396.
Kamruzzaman, M., Makino, Y., Oshita, S., 2016. Parsimonious model development for real-time monitoring of moisture in red meat using hyperspectral imaging. Food Chem. 196, 1084–1091.
Kowitz, T., Mason, R., 2003. Poor control over drying macadamia nut-in-shell on-farm causes abnormal kernel browning during roasting. In: Proceedings of the 2Nd International Macadamia Symposium. Australian Macadamia Society Lismore, pp. 55–57.
Liu, D., Sun, D.W., Zeng, X.A., 2014. Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. Food. Bioprocess Technol. 7, 307–323.

Ma, J., Jiang, X., Fan, A., Jiang, J., Yan, J., 2021. Image matching from handcrafted to deep features: A survey. Int. J. Computer vis. 129, 23–79.
Marquis Macadamias, 2024. End-To-End Processing. Marquis Macadamias, Brisbane. Available online at: < https://marquis.com/processing/our-process/> (Accessed 8 March 2024).
Maestri, D., 2023. Groundnut and tree nuts: a comprehensive review on their lipid components, phytochemicals, and nutraceutical properties. Crit. Rev. Food Sci. Nutr. 1–25.
Malmir, M., Tahmasbian, I., Xu, Z., Farrar, M.B., Bai, S.H., 2020. Prediction of macronutrients in plant leaves using chemometric analysis and wavelength selection. J. Soils Sediments 20, 249–259.
Manley, M., 2014. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. Chem. Soc. Rev. 43, 8200–8214.
Mason, R., 2000. Macadamia nut quality research: the processing challenge. Food Aust. 52, 416–419.
MathWorks Robotics System Toolbox User's Guide, 2024, (Release 2023a) https://www.mathworks.com/help/stats/fscmrmr.html (Accessed 21 May 2024).
Mehmood, T., Liland, K.H., Snipen, L., Sæbø, S., 2012. A review of variable selection methods in partial least squares regression. Chemom. Intellig. Lab. Syst. 118, 62–69.
Mohammadi-Moghaddam, T., Razavi, S.M., Taghizadeh, M., Pradhan, B., Sazgarnia, A., Shaker-Ardekani, A., 2018. Hyperspectral imaging as an effective tool for prediction the moisture content and textural characteristics of roasted pistachio kernels. J. Food Meas. Charact. 12, 1493–1502.
O'Hare, P., Stephenson, R., Quinlan, K., Vock, N., 2004. Macadamia grower's handbook. Nambour, Australia, Queensland Department of Primary Industries and Fisheries.
Panda, B.K., Mishra, G., Ramirez, W.A., Jung, H., Singh, C.B., Lee, S.-H., Lee, I., 2022. Rancidity and moisture estimation in shelled almond kernels using NIR hyperspectral imaging and chemometric analysis. J. Food Eng. 318, 110889.
Park, W., Jin, D., Kim, C.-S., 2022. Eigencontours: Novel contour descriptors based on low-rank approximation, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2667-2675.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. J. Machine Learning Res. 12, 2825–2830.
Pelliccia, D., 2018. Partial Least Squares Regression in Python, The NIRPY Research Blog NIRPY Research. Available online at: <https://nirpyresearch.com/partial-least-squares-regression-python/> (Accessed 8 March 2024).
Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Analysis Machine Intelligence 27, 1226–1238.
Ros, E., 2010. Health benefits of nut consumption. Nutrients 2, 652–682.
Rossel, R.A.V., 2008. ParLeS: Software for chemometric analysis of spectroscopic data. Chemom. Intellig. Lab. Syst. 90, 72–83.
Rossel, R.A.V., Webster, R., 2012. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database. Eur. J. Soil Sci. 63, 848–860.
Saha, D., Manickavasagan, A., 2021. Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. Curr. Res. Food Sci. 4, 28–44.
Sun, J., Zhou, X., Hu, Y., Wu, X., Zhang, X., Wang, P., 2019. Visualizing distribution of moisture content in tea leaves using optimization algorithms and NIR hyperspectral imaging. Comput. Electron. Agric. 160, 153–159.
Trueman, S.J., 2013. The reproductive biology of macadamia. Sci. Hortic. 150, 354–359.
Wallace, H.M., Walton, D., 2011. Macadamia (*Macadamia integrifolia*, *Macadamia tetraphylla* and hybrids). In: Kader, A.A., Yahia, E.M. (Eds.), Postharvest Biology and Technology of Tropical and Subtropical Fruits. Woodhead Publishing Limited, Cambridge, pp. 450–474.
Walton, D.A., Randall, B.W., Le Lagadec, M.D., Wallace, H.M., 2013. Maintaining high moisture content of macadamia nuts-in-shell during storage induces brown centres in raw kernels. J. Sci. Food Agric. 93, 2953–2958.
Walton, D., Wallace, H.M., 2008. Postharvest dropping of macadamia nut-in-shell causes damage to kernel. Postharvest Biol. Technol. 49, 140–146.
Walton, D.A., Wallace, H.M., 2011. Quality changes in macadamia kernel between harvest and farm-gate. J. Sci. Food Agric. 91, 480–484.
Walton, D., Wallace, H.M., 2012. Genetic and postharvest factors affecting macadamia kernel quality. African J. Agric. Res. 7, 2490–2495.
Wei, Y., Wu, F., Xu, J., Sha, J., Zhao, Z., He, Y., Li, X., 2019. Visual detection of the moisture content of tea leaves with hyperspectral imaging technology. J. Food Eng. 248, 89–96.
Williams, P., 2014. The RPD statistic: a tutorial note. NIR News 25, 22–23.
Williams, P., Antoniszyn, J., Manley, M., 2019. Near infrared technology: getting the best out of light. African Sun Media, Stellenbosch, South Africa.
Wu, D., Shi, H., Wang, S., He, Y., Bao, Y., Liu, K., 2012. Rapid prediction of moisture content of dehydrated prawns using online hyperspectral imaging system. Anal. Chim. Acta 726, 57–66.
Yang, J., Liu, R.H., Halim, L., 2009. Antioxidant and antiproliferative activities of common edible nut seeds. LWT-Food Sci. Technol. 42, 1–8.
Zambrano, M.V., Dutta, B., Mercer, D.G., MacLean, H.L., Touchie, M.F., 2019. Assessment of moisture content measurement methods of dried food products in small-scale operations in developing countries: A review. Trends Food Sci. Technol. 88, 484–496.
Zhu, X., Li, G., 2019. Rapid detection and visualization of slight bruise on apples using hyperspectral imaging. Int. J. Food Prop. 22, 1709–1719.